

LIBS 与变量选择 PLS 结合的含油土壤中 Cu, Ni 定量分析

朱绍农^{1, 2, 3}, 丁宇^{1, 2, 3*}, 陈雨娟^{1, 2, 3}, 邓凡^{1, 2, 3}, 陈非凡^{1, 2, 3}, 严飞^{1, 2, 3}

1. 南京信息工程大学江苏省大数据分析技术重点实验室, 江苏 南京 210044
2. 南京信息工程大学江苏省大气环境与装备技术协同创新中心, 江苏 南京 210044
3. 南京信息工程大学江苏省气象能源利用与控制工程技术研究中心, 江苏 南京 210044

摘要 土壤中重金属元素检测是环境保护事业的重点之一, 因此亟需一种能够快速检测土壤重金属浓度的定量分析手段。该研究旨在建立一种基于激光诱导击穿光谱结合偏最小二乘法的含油土壤中重金属元素快速定量检测方法。通过激光诱导击穿光谱(LIBS)获取含油土壤光谱数据, 采用偏最小二乘法(PLS)对样品中铜、镍元素进行定量分析预测, 并在此基础上, 结合区间以及后向区间法对全谱进行变量筛选, 构建形成区间偏最小二乘法(iPLS)和后向区间偏最小二乘法(BiPLS)定量分析铜、镍元素含量的模型。结果表明: 后向区间偏最小二乘法(BiPLS)在剔除了干扰信息的基础上, 保留了更多的有效光谱信息, 获得了比 PLS 和 iPLS 更好的预测结果: 铜元素的测试集预测结果的决定系数(R_p^2)和均方根误差(RMSEP)分别为 0.944 9 和 0.036 3, 相对分析误差(RPD)为 3.0; 镍元素的测试集预测结果的 R_p^2 和 RMSEP 分别为 0.933 7 和 0.041 4, RPD 为 2.6, 两元素的 BiPLS 预测结果相较于 PLS 和 iPLS 方法均有所提升。因此, 针对含油土壤重金属元素光谱信息, BiPLS 算法相较于 iPLS 和 PLS 算法更适合与 LIBS 光谱相结合, 筛选对 Cu 和 Ni 两种重金属元素定量分析贡献度较大的特征变量, 进而提升模型的预测效果。该方法将促进 LIBS 技术应用于土壤品质在线评价。

关键词 含油土壤; 重金属检测; 激光诱导击穿光谱; 偏最小二乘法

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)12-3812-06

引言

土壤中各种元素的含量是用来衡量土壤质量的重要指标之一^[1], 因此, 如何高效的掌握土壤中元素含量是农业发展的重要内容。而在工业飞速发展的大背景下, 由于工业“三废”的不合理排放, 大量的重金属污染物以不同的形式进入土壤, 其中最具代表性的便是工业用油的污染。近年来, 在石油的开采、油品的储存以及运输使用过程中, 石油泄漏事故多次发生。2004年11月18日, 陕西延安发生的特大石油泄露事故使得数百亩农田被污染; 2013年11月山东青岛的东黄输油管道原油泄漏并发生了爆炸, 不仅使得周边土壤被严重污染, 甚至造成了重大的人员伤亡。含油土壤中的重金属元素不仅会极大地破坏土壤的结构, 改变其物理化学性质, 还会影响植被的品质, 并通过食物链危害人类和动物的

生命安全, 严重威胁生态环境与人类的食品安全。因此, 检测含油土壤中的重金属浓度及其治理是目前环境保护工作的重点项目。然而使用传统方法进行土壤中重金属元素的原位分析十分困难, 加之不同重金属分析方法存在差异, 导致分析过程周期长, 成本高, 操作工序复杂, 不适合大规模使用。

激光诱导击穿光谱(laser induced breakdown spectroscopy, LIBS)是一项近年来发展极为快速的元素分析技术, 目前已经广泛应用于化工^[2]、食品^[3]、生物^[4]、考古^[5]以及农业^[6]等领域。该技术具有检测速度快、不需要样品预处理等优点^[7], 因此对于污染场地重金属元素的即时检测以及污染防控的快速反应有着极其重要的意义。但在实际应用中, 由于该项技术是通过获取等离子体发射谱线的强度信息来确定元素的含量, 而传统的基本标定法与内标定法均为单变量模型^[8], 对于自吸收效应, 基体效应等干扰无法做到有效的消除, 故需要一种准确度更高的方法来满足实际应用的需要。

收稿日期: 2019-12-18, 修订日期: 2020-04-22

基金项目: 国家自然科学基金项目(61605083), 江苏省高校自然科学研究面上项目(17KJB535002)和南京信息工程大学人才启动项目(2243141701023)资助

作者简介: 朱绍农, 1999年生, 南京信息工程大学本科生 e-mail: shaonong_zhu@nuist.edu.cn

* 通讯联系人 e-mail: dingyuaoli@163.com

在定量分析中, 偏最小二乘法 (partial least-square method, PLS)能够在自变量存在较强相关性的情况下处理光谱, 目前已经在铝合金、钢铁、煤炭等材料的 LIBS 成分检测中得到广泛应用^[9-10]。PLS 利用光谱数据作为自变量, 被分析元素浓度作为因变量, 在一定程度上可以克服传统方法的缺陷。但由于全谱建模在实验中复杂费时^[11], 且全光谱常常掺杂着非目标成分的吸收, 导致样品光谱与样本组分性质之间的关联性较差, 选取全谱作为模型的输入变量, 不仅可能会影响光谱的灵敏度, 还会影响模型的准确度。因此, PLS 结合变量选择发展出了一些改进的建模方法, 这些方法的优势也在不同领域得到了应用^[12-13]。

本研究基于 LIBS 技术对含油土壤中的铜、镍两种重金属污染元素进行分析, 建立 PLS 模型, 并将独立的 PLS 模型与区间变量选择法以及后向区间变量选择法结合, 探索对定量分析性能提升的效果。

1 方法

1.1 偏最小二乘法

PLS 算法的建模思想主要是从自变量和因变量矩阵中提取第一主成分, 并求得协方差, 再提取第二主成分, 求得协方差, 依次迭代, 最后根据交叉验证的结果, 建立最终的偏最小二乘定量回归预测分析模型。

1.2 区间偏最小二乘法

区间偏最小二乘法 (interval partial least-square method, iPLS)是由 Norgaard 等提出的波段区间选择方法^[14]。其原理主要是将数据集均分成多个同等宽度区间之后, 建立每个子区间的 PLS 模型, 再根据各个模型的交叉验证结果优选出最佳的光谱波段。

1.3 后向区间偏最小二乘法

后向区间偏最小二乘法 (backward interval partial least-square method, BiPLS)是在 iPLS 的基础上更进一步提出的算法。由于 iPLS 并不能确保选出的一个区间是最适于建模的, 故不能排除多个区间建模效果更好的可能。因此在 BiPLS 中将会进行多次计算, 并依次减少交叉验证表现最差的区间, 直到只剩下一个数据区间, 进而得出交叉验证结果最小, 即预测效果最好的波段集合。

1.4 模型的评价

以校正集与测试集的决定系数 (R_c^2 和 R_p^2)、均方根误差 (RMSEC 和 RMSEP) 和相对分析误差 (RPD) 来评价模型, 计算公式分别为

$$R^2(y, \hat{y}) = \left(\frac{\text{Cov}(y, \hat{y})}{\sqrt{\text{Var}[y]\text{Var}[\hat{y}]}} \right)^2 \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad (2)$$

$$\text{RPD} = \text{SD}/\text{RMSE} \quad (3)$$

其中 n 为校正集和测试集的样品个数, y 和 \hat{y} 分别为对应的预测值和真实值, SD 为样品的标准差。一般来说, R^2 越接近于 1, RMSE 越小, RPD 越大, 则说明模型的预测能力越好, 预测结果的准确度越高。

2 实验部分

2.1 装置

实验装置如图 1 所示。激发源采用 Q-Switched Nd : YAG 激光器 (北京镭宝, Dawa 300), 激光波长 1 064 nm, 工作频率 1 Hz, 脉冲能量设定为 150 mJ。激光器发出的高能脉冲经反射镜传递至聚焦镜 (焦距: 100 mm) 后, 聚焦在放置于三维样品台上的样品表面。烧蚀样品产生等离子体, 辐射出的光谱信号经光纤探头耦合至光谱仪。光谱仪为海洋光学 MX2500+, 波长范围 190~520 nm, 光谱分辨率 0.07 nm, 光谱积分时间设定为 1 ms。为了降低光谱信号产生前期的韧致辐射等影响, 将延时时间设定为 3 μs 。所有实验均在常温常压下完成。

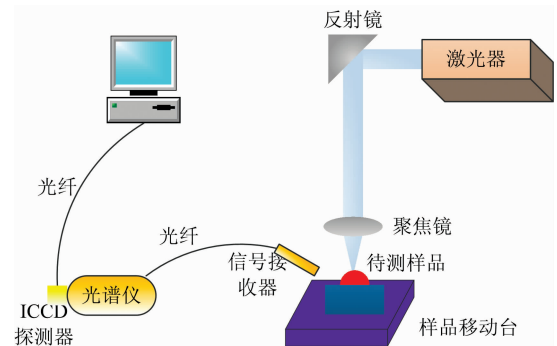


图 1 实验装置图

Fig. 1 Experimental setup

2.2 样品

实验样品中 Cu 和 Ni 的成分如表 1 所示, 由 XRF 检测获得。样品 1 是从受到油污染地区采集的原始样品, 其他样品是通过光谱纯试剂配制而成。为了降低样品不均性对实验结果的影响, 所有的样品均在 150 $^{\circ}\text{C}$ 的烤箱中烘干 6 h, 烘干后的样品研磨过 200 目筛网, 最终压制成 $\phi 20 \times 1.5$ mm 的圆片, 压力为 20 MPa。光谱信号采集过程中, 为了进一步降低激光能量波动等因素对光谱信号的影响, 每个样品采集 50 个点, 且每个点是由 5 发脉冲产生的光谱信号平均所得。建模时, 随机选取了 1, 3, 5, 6, 7, 8, 10, 11, 13, 14 和 15 号样品作为校正集, 剩余 5 个样品作为测试集。

表 1 样品中 Cu 和 Ni 含量 (%)

Table 1 The Contents list of Cu and Ni (%)

编号	Cu	Ni	编号	Cu	Ni
1	0.45	0.65	9	0.57	0.77
2	0.51	0.71	10	0.78	0.98
3	0.48	0.68	11	0.81	1.01
4	0.54	0.74	12	0.75	0.95
5	0.52	0.72	13	0.93	1.13
6	0.60	0.80	14	1.05	1.25
7	0.66	0.86	15	1.22	1.42
8	0.72	0.92	16	0.82	1.02

3 结果与讨论

3.1 定性分析

波长和强度是用于定性分析含油土壤中重金属元素的重要信息。图 2 为所有样品的光谱信号平均后得到的光谱图，结合 NIST 数据库可以发现，Cu 元素在 216.51, 324.75 和 327.39 nm 位置的特征线信号较强，Ni 元素在 338.05 和 361.49 nm 位置特征线信号较强。但由于土壤成分较为复杂，除了 Cu 和 Ni 元素，还有较多其他元素的光谱信号，如 K 和 Ca 等。这些信号的强度以及位置都会对 Cu 和 Ni 的谱线造成影响。因而探得不同的变量选择方法以剔除干扰信号，达到提升定量分析结果的目的。

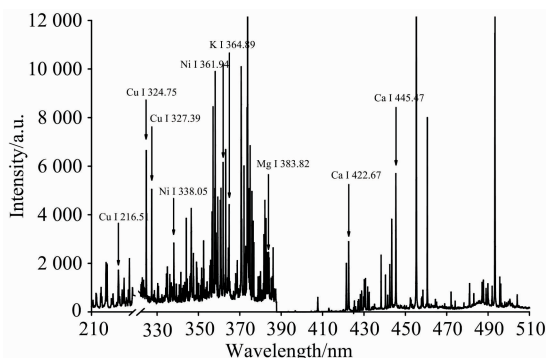


图 2 含油土壤样品典型光谱
Fig. 2 Typical spectra of soil samples

3.2 不同 PLS 模型的预测性能

3.2.1 PLS 模型预测性能

在 PLS 建模过程中，潜变量数(LV)的选择尤其重要，若 LV 太少，会导致光谱中较多信息的丢失，最终导致拟合不充分；若 LV 太多，则会导致过拟合现象，最终得到的预测误差会有显著的增大。本实验中，将通过交叉验证确定两种元素的最佳潜变量。

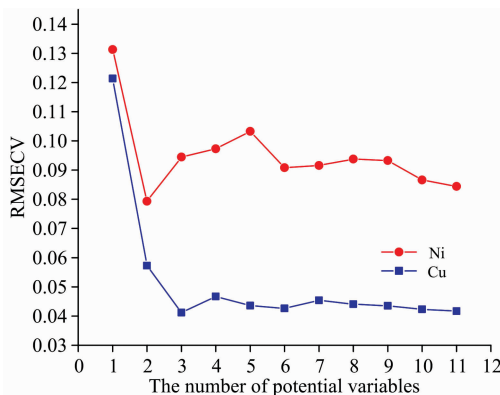


图 3 Cu 和 Ni 元素 PLS 模型下不同潜变量数对应的 RMSECV

Fig. 3 RMSECVs for PLS models with different potential variables of Cu and Ni

如图 3 所示，两种元素对应模型的 RMSECV 值先随着 LV 的增加而减小，儿后又随着 LV 的增加而有所起伏。当 LV 为 3 时，Cu 元素的 PLS 模型有最小的 RMSECV；当 LV 为 2 时，Ni 元素的 PLS 模型有最小的 RMSECV。在最佳潜变量下，建立两种重金属元素的全光谱 PLS 模型，建模预测结果如图 4、图 5 所示。

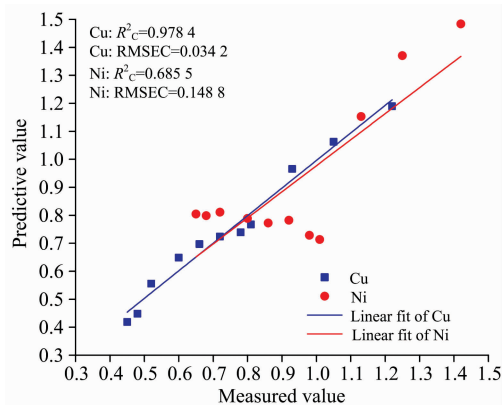


图 4 PLS 模型对校正集的预测结果

Fig. 4 Prediction results of PLS for calibration set

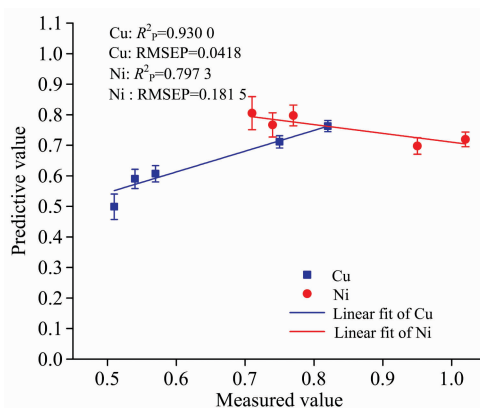


图 5 PLS 模型对测试集的预测结果

Fig. 5 Prediction results of PLS for test set

结果表明，Cu 元素的校正集 $R^2_c = 0.9784$ ， $RMSEC = 0.0342$ ；测试集预测结果的 $R^2_p = 0.9300$ ， $RMSEP = 0.0418$ 。Ni 元素的校正集预测结果的 $R^2_c = 0.6855$ ， $RMSEC = 0.1488$ ；测试集预测结果的 $R^2_p = 0.7973$ ， $RMSEP = 0.1815$ 。相较于 Cu 元素，Ni 的 PLS 模型的校正集与测试集呈现的相关性都较差，可能的原因是 Ni 元素特征谱线较弱，容易受到其他谱线的干扰。

3.2.2 iPLS 模型的预测性能

利用 iPLS 方法，将全光谱波段依次按 10~25 个区间进行等分，并在每一个区间建立 PLS 回归模型。将每次等分所获得的最小 RMSECV 作为衡量标准。如表 2 所示，在铜元素的区间划分过程中，当共划分 19 个区间时，对应区间 9 的 RMSECV 最小，故选择第 9 区间作为铜元素 iPLS 建模的输入变量。对于 Ni 元素，iPLS 筛选结果与 Cu 一致。

表 2 iPLS 模型不同区间个数对应 Cu 的 RMSECV

Table 2 RMSECVs for iPLS models with different interval numbers of Cu

区间个数	区间序号	RMSECV	变量个数
10	7	0.023 3	582
11	6	0.033 5	529
12	5	0.026 0	485
13	5	0.026 9	448
14	11	0.036 5	415
15	10	0.023 3	388
16	6	0.034 8	364
17	6	0.026 2	343
18	9	0.019 7	324
19	9	0.018 6	306
20	10	0.020 3	291
21	10	0.019 0	277
22	11	0.020 1	265
23	15	0.019 7	253
24	12	0.020 4	242
25	25	0.031 8	232

建立两元素的 iPLS 预测模型，模型的预测结果如图 6、图 7 所示。

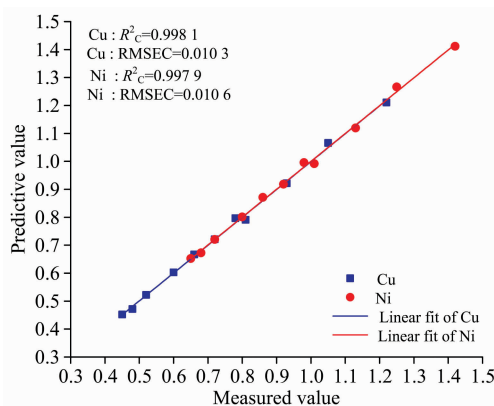


图 6 iPLS 模型对校正集的预测结果

Fig. 6 Prediction results of iPLS for calibration set

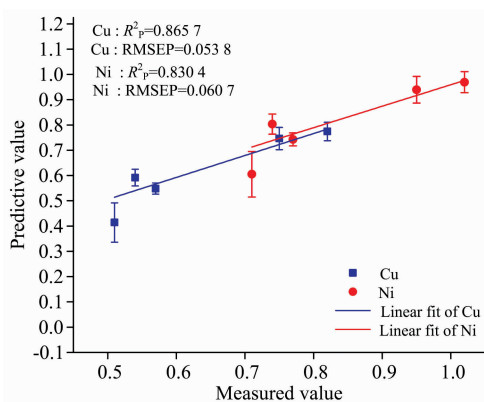


图 7 iPLS 模型对测试集的预测结果

Fig. 7 Prediction results of iPLS for test set

Cu 元素的校正集预测结果的 $R_c^2 = 0.998 1$, $RMSEC = 0.010 3$; 测试集预测结果的 $R_p^2 = 0.865 7$, $RMSEP = 0.053 8$ 。Ni 元素的校正集预测结果的 $R_c^2 = 0.997 9$, $RMSEC = 0.010 6$; 测试集预测结果的 $R_p^2 = 0.830 4$, $RMSEP = 0.060 7$ 。

3.2.3 BiPLS 模型的预测性能

虽然 iPLS 对数据集进行了一定程度的筛选，但由于其忽略了多区间建模效果更优的可能性，所以可能会丢失一些其他区间的有用信息，因此采用 BiPLS 再次进行特征变量的提取。将全谱划分为 10~25 个子区间，并挑选出最小 RMSECV 值作为建模的光谱区间集合。对于 Cu 元素，结果如表 3 所示，当所划区间数为 21 时，RMSECV 最小为 0.014 3。

表 3 BiPLS 模型下不同区间个数 Cu 的 RMSECV

Table 3 RMSECVs for BiPLS with different interval numbers of Cu

区间个数	最小 RMSECV	区间个数	变量数
10	0.023 3	1	582
11	0.020 8	5	2 645
12	0.022 1	4	1 940
13	0.022 4	5	2 238
14	0.018 4	6	2 494
15	0.025 4	6	2 327
16	0.019 8	10	3 639
17	0.019 0	8	2 739
18	0.020 6	9	2 910
19	0.018 9	9	2 756
20	0.022 1	7	2 037
21	0.014 3	7	1 940
22	0.015 4	4	1 059
23	0.015 1	10	2 530
24	0.015 8	10	2 424
25	0.016 1	9	2 094

将由表 3 确定的 21 个子区间进行联合建模，依次剔除 RMSECV 表现最差的子区间，当剔除掉 11 个子区间，剩下 10 个子区间时，RMSECV 的表现最好，10 个子区间分别为第 1, 4, 6, 8, 10, 12, 14, 15, 19 和 21 区间。对于 Ni 元素，BiPLS 筛选结果与 Cu 一致。

利用筛选出的区间分别建立两元素的 BiPLS 模型并预测，得到两元素预测的结果如图 8、图 9 所示。

Cu 元素校正集预测结果的 $R_c^2 = 0.996 3$, $RMSEC = 0.014 1$; 测试集预测结果的 $R_p^2 = 0.944 9$, $RMSEP = 0.036 3$ 。

Ni 元素校正集预测结果的 $R_c^2 = 0.994 4$, $RMSEC = 0.017 5$; 测试集预测结果的 $R_p^2 = 0.933 7$, $RMSEP = 0.041 4$ 。

3.3 模型预测性能对比

将全光谱 PLS, iPLS 以及 BiPLS 三个模型的各项结果进行比较。

铜元素建模检测结果如表 4 所示，以全光谱作为输入变

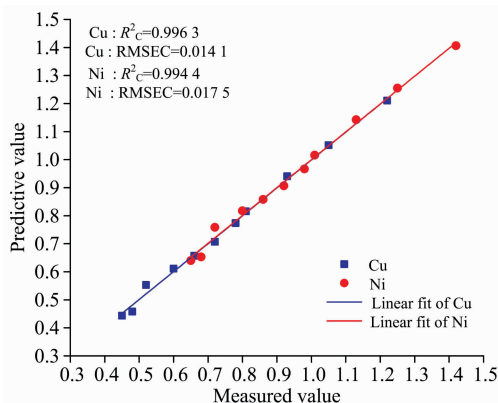


图 8 BiPLS 模型对校正集的预测结果

Fig. 8 Prediction results of BiPLS for calibration set

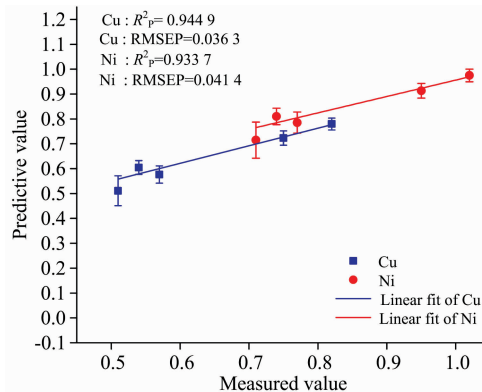


图 9 BiPLS 模型对测试集的预测结果

Fig. 9 Prediction results of BiPLS for test set

量的 PLS 模型虽然在校正集中表现良好,但测试集预测结果较差,说明光谱中掺杂了过多的干扰信息,影响了预测结果; iPLS 在筛选出建模区间后,校正集预测能力明显提高,但测试集 R_p^2 低于 PLS, RMSEP 的结果也高于 PLS,可能由于 iPLS 模型剔除了部分的有用信息,构建的模型准确度有所降低,预测能力变差; BiPLS 模型在 iPLS 的基础上保留了更多的有用区间,能够利用的光谱信息更多。测试集预测结果为三种模型中最佳, R_p^2 由原来的 0.930 0 提高到了 0.944 9, RMSEP 由原来的 0.041 8 减小到了 0.036 3, RPD 由 2.5 提升至 3.0, 数据表明 BiPLS 模型的准确度更佳,更适用于检测土壤中的铜元素含量。

镍元素建模检测结果如表 5 所示,以全光谱建模的 PLS 模型预测能力较差,校正集和测试集的预测结果均为三种模型中最差, $RPD < 1$, 可能对于镍元素来说,光谱中的干扰信息过多,全光谱建模不适用于检测其浓度; iPLS 模型在筛选出建模区间后,校正集和测试集的预测结果表现均有明显提高,说明所选区间的代表性较好,光谱中的冗余信息也被有效的剔除;而 BiPLS 在 iPLS 的基础上,保留了更多的光谱信息,故其测试集表现更好。由实验数据可知, R_p^2 由原来的 0.797 3 提高到了 0.933 7, RMSEP 由原来的 0.181 5 减小到了 0.041 4, 预测结果更接近于真实值, RPD 由 0.3 提升至 2.6。综合来看, BiPLS 模型的准确度与预测能力更好,更适用于检测土壤中的镍元素含量。

表 4 Cu 元素的 PLS, iPLS, BiPLS 模型结果比较

Table 4 Comparison of PLS, iPLS and BiPLS models for Cu element

算法类别	变量数量	校正集		测试集		
		R_c^2	RMSEC	R_p^2	RMSEP	RPD
PLS	5 819	0.978 4	0.034 2	0.930 0	0.041 8	2.5
iPLS	253	0.998 1	0.010 3	0.865 7	0.053 8	2.8
BiPLS	1 940	0.996 3	0.014 1	0.944 9	0.036 3	3.0

表 5 Ni 元素的 PLS, iPLS, BiPLS 模型结果比较

Table 5 Comparison of PLS, iPLS and BiPLS models for Ni element

算法类别	变量数量	校正集		测试集		
		R_c^2	RMSEC	R_p^2	RMSEP	RPD
PLS	5 819	0.685 5	0.148 8	0.797 3	0.181 5	0.3
iPLS	253	0.997 9	0.010 6	0.830 4	0.060 7	2.5
BiPLS	1 940	0.994 4	0.017 5	0.933 7	0.041 4	2.6

4 结论

分别采用了 PLS, iPLS 和 BiPLS 对土壤中的 Cu、Ni 两种元素的含量进行了建模,并在建模过程中进行了不同方式的特征波长区间的筛选。实验结果表明: PLS 由于采用全光谱建模,光谱中过多的冗余信息干扰了建模结果,所建模型

的准确度与预测能力均较差; iPLS 虽然筛选出了部分波段,建模结果的表现有所上升,但由于其剔除的有用信息过多,导致模型的预测能力和准确度没有明显提升; BiPLS 在剔除了干扰信息的基础上,保留了更多的光谱信息,模型的预测能力有明显提高,且预测结果的准确性最佳,铜元素测试集预测结果的 R_p^2 和 RMSEP 分别为 0.994 9 和 0.036 3, RPD 达到 3.0; 镍元素测试集预测结果的 R_p^2 和 RMSEP 分别为

0.933 7 和 0.041 4, RPD 达到 2.6, 两元素的预测结果较其他两种方法均有提升。实验表明, LIBS 结合 BiPLS 能够快速准确的对土壤中的重金属元素进行定量检测, 该方法可以为

含油土壤的重金属元素检测以及环境保护领域的其他指标分析提供一定的技术参考。

References

- [1] Croft H, Kuhn N J, Anderson K. *Catena*, 2012, 94: 64.
- [2] Qiu Y, Wu J, Li X W, et al. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2018, 149: 48.
- [3] Abdel-Salam Z, Alexeree Sh M I, Harith M A. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2018, 149: 112.
- [4] Viljanen J, Zhao H, Zhang Z, et al. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2018, 149: 76.
- [5] Senesi G S, Manzini D, De Pascale O. *Applied Geochemistry*, 2018, 96: 87.
- [6] Sha W, Niu P, Zhen C, et al. *Journal of Applied Spectroscopy*, 2018, 85(4): 653.
- [7] Singh J, Kumar R, Awasthi S, et al. *Food Chemistry*, 2017, 221: 1778.
- [8] LU Cui-ping, LIU Wen-qing, ZHAO Nan-jing, et al(鲁翠萍, 刘文清, 赵南京, 等). *Acta Physica Sinica(物理学报)*, 2011, 60(4): 388.
- [9] Dyar M D, Fassett C I, Giguere S, et al. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2016, 123: 93.
- [10] Mei Y G, Cheng S S, Hao Z Q, et al. *Plasma Science and Technology*, 2019, 21(3): 171.
- [11] Sorol N, Arancibia E, Bortolato S A, et al. *Chemometrics and Intelligent Laboratory Systems*, 2010, 102(2): 100.
- [12] Ma H L, Wang J W, Chen Y J, et al. *Food Chemistry*, 2017, 215: 108.
- [13] Zhou S, Wang Z, Lu L, et al. *Infrared Physics & Technology*, 2016, 77: 1.
- [14] Norgaard L, Hahn M T, Knudsen L B, et al. *International Dairy Journal*, 2005, 15(12): 1261.

Quantitative Analysis of Cu and Ni in Oil-Contaminated Soil by LIBS Combined With Variable Selection Method and PLS

ZHU Shao-nong^{1,2,3}, DING Yu^{1,2,3*}, CHEN Yu-juan^{1,2,3}, DENG Fan^{1,2,3}, CHEN Fei-fan^{1,2,3}, YAN Fei^{1,2,3}

1. Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science & Technology, Nanjing 210044, China
2. Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, Nanjing University of Information Science & Technology, Nanjing 210044, China
3. Jiangsu Engineering Research Center on Meteorological Energy Using and Control, Nanjing University of Information Science & Technology, Nanjing 210044, China

Abstract The detection of heavy metals in soil is one of the emphases of environmental protection. This paper aims to establish a fast and quantitative method for the determination of heavy metal elements in soil, based on LIBS and combined with the PLS method. We used PLS model to quantitatively analyze and predict the contents of Cu and Ni elements in oil-contaminated samples. On this basis, the variables of the full spectrum were screened by combining the Interval method and the Backward Interval method, which formed the Interval Partial Least Square (iPLS) and the Backward Interval Partial Least Square (BiPLS). The experimental results showed that the BiPLS method retained more spectral information after removing the interference information, and obtained better-predicted results than PLS and iPLS. The R²P and RMSEP of the predicted results of the test set for the copper element are 0.944 9 and 0.036 3, respectively, and the RPD reached 3.0. Those of the predicted results of the test set for nickel element are 0.933 7 and 0.041 4, respectively, and the RPD reached 2.6. Compared with the PLS and iPLS methods, the prediction results of the BiPLS method of the two elements were significantly optimized, the predictive ability was significantly improved, and the accuracy was much better. Therefore, In the analysis of heavy metal elements in oil-contaminated soil by LIBS technique, BiPLS is more suitable than iPLS and PLS for screening the feature variables that contribute greatly to the quantitative analysis of Cu and Ni elements, so as to improve the prediction effect. This method will promote the application of LIBS technology to the online evaluation of soil quality.

Keywords Oil-contaminated soil; Heavy metal detection; LIBS; Partial least squares

* Corresponding author

(Received Dec. 18, 2019; accepted Apr. 22, 2020)